# A Graphical Characterization of Lattice Conditional Independence Models†

Steen A. Andersson, David Madigan, Michael D. Perlman*, and Christopher M. Triggs

Indiana University, University of Washington, and University of Auckland

**Abstract**

Lattice conditional independence (LCI) models for multivariate normal data recently have been introduced for the analysis of non-monotone missing data patterns and of nonnested dependent linear regression models (≡ seemingly unrelated regressions). It is shown here that the class of LCI models coincides with a subclass of the class of graphical Markov models determined by acyclic digraphs (ADGs), namely, the subclass of *transitive* ADG models. An explicit graph-theoretic characterization of those ADGs that are Markov equivalent to some transitive ADG is obtained. This characterization allows one to determine whether a *specific* ADG $D$ is Markov equivalent to some transitive ADG, hence to some LCI model, in polynomial time, without an exhaustive search of the (possibly superexponentially large) equivalence class $[D]$. These results do *not* require the existence or positivity of joint densities.

## 1. Introduction: Graphical Markov Models.

The use of directed graphs to represent possible dependencies among statistical variables dates back to the work of Sewall Wright (1921) in population genetics and has since generated considerable research activity in the social, natural, engineering, and computational sciences. Both directed and undirected graphs have found extensive applications: the former as "path diagrams" for structural equation models in psychometrics and econometrics, as "Bayesian networks" for expert systems in artificial intelligence, and as "influence diagrams" in operations research and management science, the latter as models of dependence for spatial stochastic processes, image analysis, and contingency tables. The books by Pearl (1988), Whittaker [W] (1990), Spirtes, Glymour, and Scheines (1993), Edwards (1995), Cox and Wermuth (1996), Lauritzen [L] (1996), and Jensen (1996) nicely describe these developments.

Since 1980, particular attention has been directed to graphical Markov models determined by conditional independence relations among the variables, i.e., by the Markov properties specified by the graph (cf. Figure 1.1). Graphical Markov models

†Research supported in part by the U. S. National Science Foundation and National Security Agency.

**Address for correspondence*: Department of Statistics, Box 354322, University of Washington, Seattle, WA 98195-4322, USA. (michael@stat.washington.edu)

determined by acyclic directed graphs (ADGs), formally defined in Section 2, admit especially simple statistical analysis. In particular, ADG models admit convenient recursive factorizations of their joint probability density functions (pdf) (Lauritzen *et al* (1990)), provide an elegant framework for Bayesian analysis (Spiegelhalter and Lauritzen (1990)), and, in expert system applications, allow simple causal interpretations (Lauritzen and Spiegelhalter (1988)). In the multivariate normal and multinomial cases, the likelihood function (LF) (i.e., both the joint probability density function and the parameter space) factorizes recursively, which yields explicit maximum likelihood estimates (MLE) and likelihood ratio tests (LRT) - cf. [W] (1990), [L] (1996), and Andersson and Perlman [AP] (1996). The only undirected graphical (UG) models that provide these conveniences are the *decomposable* models, those UG models that have the same Markov properties as ADG models (Wermuth and Lauritzen (1993), Dawid and Lauritzen (1993), Andersson, Madigan, and Perlman [AMP] (1996a)).

$$b \rightarrow a \rightarrow c \qquad b \leftarrow a \rightarrow c \qquad b \leftarrow a \leftarrow c \qquad b \rightarrow a \leftarrow c$$

Figure 1.1. Four acyclic digraphs with the vertex set $V \equiv \{a, b, c\}$. The first three graphs are Markov equivalent: each specifies the single conditional independence $X_b \perp X_c \mid X_a$ abbreviated as $b \perp c \mid a$, while the joint pdf factors as $f(a, b, c) = f(b)f(a \mid b)f(c \mid a) \equiv f(b \mid a)f(a)f(c \mid a) \equiv f(b \mid a)f(a \mid c)f(c)$. The fourth graph specifies the independence $b \perp c$ and the factorization $f(b)f(a \mid b, c)f(c)$.

For these reasons, ADG models have become popular across a wide range of applications; see, for example, Lauritzen and Spiegelhalter (1988), Pearl (1988), Neapolitan (1990), Spiegelhalter and Lauritzen (1990), Spiegelhalter *et al* (1993), Madigan and Raftery (1994), and York *et al* (1995). Indeed, the active "Uncertainty in Artificial Intelligence" community focuses much of its effort on ADG models.

Lattice conditional independence (LCI) models were introduced by [AP] (1988, 1993) under the additional assumption of multivariate normality, motivated by analogy with the lattice structure of balanced ANOVA designs (cf. Andersson (1990)). In an LCI model, formally defined in Sections 4 and 7, conditional independence relations among the variables are specified by the intersection properties of a finite distributive lattice (cf. Figure 1.2). LCI models share the desirable statistical properties of ADG models: the LF factors recursively, again yielding explicit MLEs and LRTs in the multinomial and multivariate normal cases ([AP] (1993, 1995a, 1995b), Perlman and Wu [PW] (1996)).

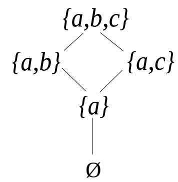



Figure 1.2. A finite distributive lattice, represented as a ring $\boldsymbol{K}$ of subsets of $\{a, b, c\} \equiv I$ (cf. §7). The LCI model $\mathbf{L}(\boldsymbol{K})$ specifies the conditional independence $\{a,b\}\perp\{a,c\} \mid \{a\}$, or equivalently, $b\perp c \mid a$, and the joint pdf factors as $f(a, b, c) = f(b \mid a)f(a)f(c \mid a)$.

In the multivariate normal and multinomial cases, LCI models are precisely suited for the analysis of non-monotone missing data patterns ([AP] (1991), [PW] (1996)). Suppose, for example, that $X_1,\ldots,X_n$ is a sample from a $p$-variate normal distribution with unknown mean and covariance, and that for each $j = 1,\ldots,n$ only a subvector $(X_j)_{K_j}$ is observed, where, for $K \subseteq \{1,\ldots,p\}$, $X_K := (X_i \mid i \in K)$. The set $\boldsymbol{S} :=(K_j \mid j = 1,\ldots,n)$ is the *observed data pattern*. It is well known that if $\boldsymbol{S}$ is *monotone*, i.e., totally ordered under inclusion, then the likelihood function (LF) based on the observed data factors recursively into a product of conditional normal LFs, allowing explicit likelihood analysis by standard linear methods (Anderson (1957)). If $\boldsymbol{S}$ is non-monotone, however, no such factorization exists. In this case, if we let $\boldsymbol{K} \equiv \boldsymbol{K}(\boldsymbol{S})$ be the ring of subsets of $\{1,\ldots,p\}$ generated by $\boldsymbol{S}$ and impose the parsimonious set of CI constraints determined by the LCI model $\mathbf{L}(\boldsymbol{K})$ (§7), then the joint LF does factor recursively and explicit likelihood analysis by standard linear methods is again possible. (For example, if $p = 3$ and $\boldsymbol{S} = \{\{1,2\}, \{1,3\}, \{1,2,3\}\}$, then $\boldsymbol{K} = \{ø, \{1\}, \{1,2\}, \{1,3\}, \{1,2,3\}\}$ has the form shown in Figure 1.2 and the CI constraint specified by $\mathbf{L}(\boldsymbol{K})$ is just $2\perp 3 \mid 1$.)

Similarly, LCI models are precisely suited for the analysis of a family of *non-nested, dependent* univariate regression models ([AP] (1994)), which includes Zellner's (1962) well-known *seemingly unrelated regressions* model as a special case. Analogously to the missing data problem, if the family $\{U_1,\ldots, U_p\}$ of regression subspaces is nested, then under the assumption of normality the joint LF factors recursively and standard linear regression methods apply. No such recursive factorization exists in the non-nested case, but the poset formed by the regression subspaces $\{U_1,\ldots, U_p\}$ under inclusion determines (see §7) a ring $\boldsymbol{K}$ of subsets of $\{1,\ldots, p\}$ such that if the CI constraints determined by the LCI model $\mathbf{L}(\boldsymbol{K})$ are imposed, then the LF factors recursively and admits an explicit likelihood analysis. Furthermore, this set of CI constraints is parsimonious (minimal) not only with respect to the class of all LCI models that allow a recursive factorization for the combined regression problem *but also with respect to the* *larger (§4) class of all ADG models that allow such a recursive factorization* ([AP] (1996)). In this sense, LCI models form a distinguished subclass (§4) of the class of ADG models.

These considerations have raised the following question: what is the exact relation between the classes of LCI and ADG models? AMP and Triggs [AMPT] (1995) showed that if consideration is restricted to multivariate distributions with *positive joint densities*, then the class of LCI models coincides with a subclass of the class of ADG models, namely, the subclass determined by all *transitive* ADGs (see Section 3). The first purpose of the present paper is to show that this restriction on the densities can be dropped; in fact, not even the existence of joint densities need be assumed. This result, developed in Sections 2 - 4, appears as Theorem 4.1.

Next we consider the question of whether a *specific* ADG $D$ determines a statistical model that coincides with some LCI model, i.e., whether $D$ is Markov equivalent to *some* LCI model. One may apply a standard polynomial-time algorithm[1] to determine whether $D$ is transitive; if so, then the answer is *yes*, by Theorem 4.1. If $D$ is not transitive, however, the answer is not necessarily *no*, for the following reason. Whereas at most one undirected graph can be associated with a given graphical Markov model, there may be several (often many) ADGs that determine the same Markov model - see Figure 1.1 and Section 5 for examples. The family of all ADGs with a given set of vertices is naturally partitioned into *Markov-equivalence classes*, each class being associated with a unique statistical model. Although $D$ may not be transitive, its Markov-equivalence class may contain a transitive ADG, in which case $D$ *is* Markov equivalent to some LCI model.

Thus, in order to determine whether a specific ADG $D$ is Markov equivalent to some LCI model, one must answer the following more complex question: *does the Markov-equivalence class* $[D]$ *contain at least one transitive ADG*? This can be decided by exhaustive search of all members of $[D]$ (see Section 5), but since $[D]$ can be super-exponentially large, exhaustive search of $[D]$ is computationally unfeasible for large graphs. The second main purpose of this paper is to provide a computationally feasible characterization, based on the *essential graph* $D^*$, of those $D$ such that $[D]$ contains at least one transitive ADG and therefore is Markov equivalent to some LCI model.

[AMP] (1996b) show that for each ADG $D$, the equivalence class $[D]$ can be uniquely represented by a certain Markov-equivalent *chain graph*[2] $D^*$, the *essential graph*[3]

---

[1]Such algorithms require at most $O(n^3)$ operations, where $n$ is the number of vertices.

[2]Chain graphs (= adicyclic graphs) may have both directed and undirected edges but may contain no (partially) directed cycles; they include both ADGs and UGs as special cases.

associated with the equivalence class. They present an explicit characterization of those graphs $G$ such that $G = D^*$ for some ADG $D$, then apply this characterization to obtain a polynomial-time algorithm[4] for constructing $D^*$ from $D$. These results are reviewed in Section 5 of the present paper. In Section 6, we characterize those essential graphs $D^*$ that are Markov equivalent to some *transitive* ADG and show that this characterization can be verified in polynomial time. Combined with the algorithm for constructing $D^*$ from $D$, this establishes the computational feasibility of determining whether a specific ADG model is Markov equivalent to some LCI model. Also, we present an orientation algorithm that produces a Markov-equivalent transitive ADG from $D^*$ if one exists.

In the statistical applications of LCI models reviewed above, the finite distributive lattice determining the LCI model is presented as a ring of subsets. In Section 7 we relate this representation to the general definition of an LCI model given in Section 4.

Some basic definitions, and terminology concerning graphs are summarized in the Appendix, which the reader is invited to review first.

The representation of an LCI model as a transitive ADG model was first suggested by Steffen Lauritzen, whom we thank for many helpful suggestions.

## 2. Graphical Markov Models Determined by Acyclic Digraphs.

The initial discussion in this section follows Lauritzen *et al.* (1990) and [AP] (1996). We consider multivariate probability distributions $P$ on a product probability space $\boldsymbol{X} \equiv \times(\boldsymbol{X}_a \,|\, a \in V)$, where $V$ is a finite index set and each $\boldsymbol{X}_a$ is sufficiently regular to ensure the existence of regular conditional probability distributions. Such a distribution is conveniently represented by a random variate $X := (X_a \,|\, a \in V) \in \boldsymbol{X}$. For any subset $A \subseteq V$, we define $X_A := (X_a \,|\, a \in A)$. Often we abbreviate $X_a$ and $X_A$ by $a$ and $A$, respectively, and define $X_\emptyset \equiv$ constant. For three pairwise disjoint subsets $A$, $B$, and $C$ of $V$, we write $A \perp B \,|\, C\ [P]$ if $X_A$ and $X_B$ are conditionally independent given $X_C$ under $P$. If $A$, $B$, and $C$ are not disjoint, then $A \perp B \,|\, C\ [P]$ is defined to mean $[A \setminus (B \cup C)] \perp [B \setminus (A \cup C)] \,|\, C\ [P]$.

A graphical Markov model is defined by a collection of conditional independencies among the component random variates $(X_a \,|\, a \in V)$, which collection is represented by an acyclic directed graph (ADG) $D \equiv (V, E)$ with vertex set $V$ in the following (equivalent) ways:

---

[3]The essential graph associated with an (equivalence class of) ADG(s) was first introduced by Verma and Pearl (1990) as the *completed pattern* associated with the ADG.

[4]Chickering (1995) and Meek (1995) also have obtained polynomial-time algorithms for this construction.

**Definition 2.1.** Let $D$ be an ADG. A probability measure $P$ on $\boldsymbol{X}$ is said to satisfy:

(i) the *local Markov property* (LMP) relative to $D$ if, for every $a \in V$,

(2.1) $$a \perp (\mathrm{nd}(a) \setminus \mathrm{pa}(a)) \mid \mathrm{pa}(a)\ [P];$$

(ii) the *global Markov property* (GMP) relative to $D$ if

(2.2) $$A \perp B \mid S\ [P]$$

whenever $S$ separates $A$ and $B$ in $(G_{\mathrm{an}(A \cup B \cup S)})^{\mathrm{m}}$;

(iii) the *well-numbered Markov property* (WNMP) relative to $D$ if, for each $k = 2, ..., n$,

(2.3) $$a_k \perp (\{a_1, ..., a_{k\text{-}1}\} \setminus \mathrm{pa}(a_k)) \mid \mathrm{pa}(a_k)\ [P],$$

where $n = |V|$ and $a_1, ..., a_n$ is a *well-numbering* of the members of $V$, i.e., $r < s \Rightarrow a_r \in \mathrm{nd}(a_s)$. Since $a_k \notin \mathrm{pa}(a_k)$ and $\mathrm{pa}(a_k) \subseteq \{a_1, ..., a_{k\text{-}1}\}$, (2.3) is equivalent to

(2.4) $$(\mathrm{pa}(a_k) \cup \{a_k\}) \perp \{a_1, ..., a_{k\text{-}1}\} \mid \mathrm{pa}(a_k)\ [P].$$

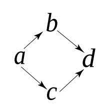



Figure 2.1. An ADG $D$ with vertex set $V \equiv \{a, b, c, d\}$. The local and global Markov properties specify the two conditional independences $b \perp c \mid a$ and $a \perp d \mid b, c$.

The term *well-numbering* appears as *never-decreasing* in [AP] (1993). The following result combines Propositions 4 and 5 of Lauritzen *et al* (1990):

**Theorem 2.1.** Let $D$ be an ADG. For any probability distribution $P$ on $\boldsymbol{X}$, GMP $\Leftrightarrow$ LMP $\Leftrightarrow$ WNMP.

It follows from Theorem 2.1 that the WNMP does not depend on the well-numbering chosen for $V$.

**Definition 2.2.** Let $D$ be an ADG. The set $\mathbf{M}_{\boldsymbol{X}}(D)$ of all probability distributions on $\boldsymbol{X}$ that satisfy the three equivalent Markov properties LMP, GMP, and WNMP relative to $D$ is called the *Markov model* determined by the ADG $D$, or, simply, the *ADG model* determined by $D$.

In applications an additional parametric assumption, such as multivariate normality, is often imposed on ADG models, but we shall not do so here. Many examples of specific ADG models appear in the references cited in Section 1. It is again emphasized that different ADGs can determine the same Markov model.

In order to relate ADG and LCI models in Section 4, we now introduce a new Markov-type property determined by an ADG $D \equiv (V, E)$. A subset $A \subseteq V$ is called *ancestral* in $D$ if $a \in A$ whenever $a \in V$, $b \in A$, and $a \leq b$ in $D$. The *ancestral ring* $\mathbf{A}(D)$ is defined to be the collection of all ancestral subsets of $D$. Clearly, $\mathbf{A}(D)$ is a ring of subsets of $V$, i.e., $\mathbf{A}(D)$ is closed under unions and intersections, hence $\mathbf{A}(D)$ is a finite distributive lattice under these set operations, and $\emptyset, V \in \mathbf{A}(D)$. For any subset $A \subseteq V$, $\mathrm{an}(A)$ denotes the smallest ancestral set containing $A$: $\mathrm{an}(A) = \{b \mid b \leq a \text{ for some } a \in A\}$.

**Definition 2.3.** Let $D$ be an ADG. A probability distribution $P$ on $\boldsymbol{X}$ is said to satisfy the *lattice conditional independence property* (LCIP) relative to $D$ if, for every pair $A, B \in \mathbf{A}(D)$,

(2.5) $$A \perp B \mid A \cap B,$$

or, equivalently,

(2.6) $$(A \setminus B) \perp (B \setminus A) \mid A \cap B.$$

**Definition 2.4.** Let $D$ be an ADG. The set $\mathbf{L}_{\boldsymbol{X}}(D)$ of all probability distributions on $\boldsymbol{X}$ that satisfy the LCIP relative to $D$ is called the *lattice conditional independence model* (LCI model) determined by $D$.

**Theorem 2.2.** Let $D$ be an ADG. For any probability distribution $P$ on $\boldsymbol{X}$, GMP $\Rightarrow$ LCIP. Thus, $\mathbf{M}_{\boldsymbol{X}}(D) \subseteq \mathbf{L}_{\boldsymbol{X}}(D)$.

**Proof.** For any pair $A, B \in \mathbf{A}(D)$, the GMP will imply (2.6) provided that $(A \setminus B)$ and $(B \setminus A)$ are separated by $A \cap B$ in

$$(D_{\mathrm{an}((A \setminus B) \cup (B \setminus A) \cup (A \cap B))})^{\mathrm{m}} = (D_{\mathrm{an}(A \cup B)})^{\mathrm{m}} = (D_{A \cup B})^{\mathrm{m}};$$

the second equality follows since $A \cup B \in \boldsymbol{A}(D)$. To establish this separation, consider a pair $a \in A \setminus B$ and $b \in B \setminus A$ such that there exists a path $\{a \equiv c_1, \ldots, c_n \equiv b\}$ in $(D_{A \cup B})^{\mathrm{m}}$. Then there must exist an adjacent pair $c_k, c_{k+1}$ such that $c_k \in A \setminus B$ and $c_{k+1} \in B$. By the definition of the moral graph, this adjacency occurs iff either (i) $c_k \to c_{k+1} \in D_{A \cup B}$, (ii) $c_k \leftarrow c_{k+1} \in D_{A \cup B}$, or (iii) $c_k \to d$ and $c_{k+1} \to d$ in $D_{A \cup B}$ for some $d \in D_{A \cup B}$. In case (i), $c_k \in B$

since $B$ is ancestral, hence $c_k \in A \cap B$; similarly, $c_{k+1} \in A \cap B$ in case (ii). In case (iii), either $d \in A$ or $d \in B$, implying that $c_{k+1} \in A \cap B$ or $c_k \in A \cap B$, respectively. In all cases, therefore, the path between $a$ and $b$ must pass through $A \cap B$, hence $A \cap B$ separates $A \setminus B$ and $B \setminus A$ in $(D_{A \cup B})^m$. This completes the proof.

**Remark 2.1.** For nonsingular multivariate normal distributions, Theorem 2.2 was established (in a somewhat different but equivalent form) in Theorem 5.1 of [AMPT] (1995). Theorem 2.2 was discovered independently by Koster (1996), Proposition 3.2.

## 3. Transitive Acyclic Digraphs.

An ADG $D \equiv (V, E)$ is said to be a *transitive ADG* ($\equiv$ TADG) if $a \to b \in D$ and $b \to c \in D \Rightarrow a \to c \in D$, where $a, b, c \in V$. Equivalently, $D$ is transitive if $\mathrm{an}(a) \setminus \{a\} = \mathrm{pa}(a)$ for all $a \in V$.

An ADG $D \equiv (V, E)$ becomes a partially ordered set ($\equiv$ poset) $(V, \leq)$ under the partial ordering $\leq$ defined by $a \leq b$ if $a \in \mathrm{an}(b)$. Under this partial ordering, $a < b$ iff $a \in \mathrm{an}(b) \setminus \{b\}$, i.e., iff there is a directed path from $a$ to $b$ in $D$. Thus, $D$ is transitive iff $a < b \Rightarrow a \to b \in D$, in which case the two relations are equivalent: $a < b \Leftrightarrow a \to b \in D$. Therefore, finite posets and TADGs are identical mathematical objects.

**Theorem 3.1.** Let $D$ be a TADG. For any probability distribution $P$ on $\boldsymbol{X}$, LCIP $\Rightarrow$ WNMP. Thus, for a TADG, GMP $\Leftrightarrow$ LMP $\Leftrightarrow$ WNMP $\Leftrightarrow$ LCIP and $\mathbf{M}_{\boldsymbol{X}}(D) = \mathbf{L}_{\boldsymbol{X}}(D)$.

**Proof.** Assume that $P$ satisfies the LCIP relative to $D$. To verify (2.4), use the transitivity of $D$ to rewrite (2.4) as

$$\mathrm{an}(a_k) \perp \{a_1, ..., a_{k-1}\} \mid \mathrm{an}(a_k) \setminus \{a_k\} \ [P]. \quad (3.1)$$

But $\mathrm{an}(a_k)$ and $\{a_1, ..., a_{k-1}\}$ are ancestral sets whose intersection is $\mathrm{an}(a_k) \setminus \{a_k\}$, so (3.1) is implied by the LCIP. The second statement then follows from Theorems 2.1 and 2.2.

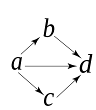



Figure 3.1. A transitive ADG $D$ with vertex set $V \equiv \{a, b, c, d\}$. The model $\mathbf{M}_{\boldsymbol{X}}(D) \equiv \mathbf{L}_{\boldsymbol{X}}(D)$ specifies the single conditional independence $b \perp c \mid a$. Note that $\mathbf{A}(D) \equiv \{\emptyset, \{a\}, \{a, b\}, \{a, c\}, \{a, b, c\}, \{a, b, c, d\}\}$ is isomorphic to the lattice $L$ in Figure 4.1 - compare to (4.1).

## 4. Lattice Conditional Independence Models.

The general class of LCI models is now defined and shown to coincide with the class of all graphical Markov models determined by TADGs. We refer to Davey and Priestley (1990), Ch. 8, for a review of the fundamental duality between finite distributive lattices and finite posets.

Let $L \equiv L(\vee, \wedge)$ be a finite distributive lattice with minimum element 0. The subset $\mathbf{J}(L)$ of *join-irreducible elements* of $L$ is defined as follows:

$$\mathbf{J}(L) := \{j \in L \mid j \neq 0, j = k \vee l \Rightarrow j = k \text{ or } j = l\}.$$

Then $(\mathbf{J}(L), \leq)$ becomes a finite poset under the partial ordering $\leq$ inherited from $L$: $j \leq k$ iff $j \wedge k = j$.

A subset $A \subseteq \mathbf{J}(L)$ is *ancestral* if $j \in A$ whenever $j \in \mathbf{J}(L)$, $k \in A$, and $j \leq k$. Note that the ring $\mathbf{A}((\mathbf{J}(L), \leq))$ of all ancestral subsets of the poset $(\mathbf{J}(L), \leq)$ is identical to the ancestral ring $\mathbf{A}((\mathbf{J}(L), E^{<}))$, where $(\mathbf{J}(L), E^{<})$ is the TADG given by $E^{<} := \{(j,k) \in \mathbf{J}(L) \times \mathbf{J}(L) \mid j < k\}$. This justifies the use of the same notation "an( )" for posets (see (4.1)) and for ADGs.

Birkhoff's Theorem (cf. Davey and Priestley (1990), §8.17) states that the mapping

(4.1)
$$L \to \mathbf{A}((\mathbf{J}(L), \leq))$$
$$l \to \mathrm{an}(l) := \{j \in \mathbf{J}(L) \mid j \leq l\}$$

determines a lattice isomorphism between the finite distributive lattice $L$ and the ring $\mathbf{A}((\mathbf{J}(L), \leq))$. This implies, in particular, that for every pair $l, m \in L$,

(4.2)
$$\mathrm{an}(l \wedge m) = \mathrm{an}(l) \cap \mathrm{an}(m).$$

When convenient, we shall identify $\mathrm{an}(l)$ with $l$ for $l \in L$.

A general LCI model determined by the finite distributive lattice $L$ consists of a family of multivariate probability distributions $P$ on a product probability space $\boldsymbol{X} \equiv \times(\boldsymbol{X}_j \mid j \in \mathbf{J}(L))$ indexed by $\mathbf{J}(L)$, where again each $\boldsymbol{X}_j$ is sufficiently regular to ensure the existence of regular conditional probability distributions. It is again convenient to represent such a distribution by a random variate $X := (X_j \mid j \in \mathbf{J}(L)) \in \boldsymbol{X}$. For any subset $J \subseteq \mathbf{J}(L)$, define $X_J := (X_j \mid j \in J)$, abbreviate $X_j$ and $X_J$ by $j$ and $J$, respectively, and define $X_\emptyset \equiv$ constant.

**Definition 4.1.** Let $L$ be a finite distributive lattice. A probability measure $P$ on $\boldsymbol{X}$ is said to satisfy the *lattice conditional independence property* (LCIP) relative to $L$ if, for every $l, m \in L$,

$$\text{an}(l) \perp \text{an}(m) \mid \text{an}(l \wedge m)\ [P]; \tag{4.3}$$

by identifying an($l$) with $l$, (4.3) can be expressed in the simpler form

$$l \perp m \mid l \wedge m\ [P]. \tag{4.4}$$

**Definition 4.2.** Let $L$ be a finite distributive lattice. The set $\mathbf{L}_{\boldsymbol{X}}(L)$ of all probability distributions on $\boldsymbol{X}$ that satisfy the LCIP relative to $L$ is called the *lattice conditional independence (LCI) model* determined by $L$.

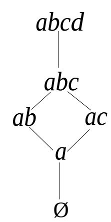



Figure 4.1. A finite distributive lattice $L$. The LCI model $\mathbf{L}_{\boldsymbol{X}}(L)$ specifies the conditional independence $ab \perp ac \mid a$. Here $\mathbf{J}(L) = \{a, ab, ac, abcd\}$ and the TADG $(\mathbf{J}(L), E^{<})$ is isomorphic to the TADG $D$ in Figure 3.1.

**Theorem 4.1.** The class of LCI models coincides with the class of TADG models.

**Proof.** For a general LCI model $\mathbf{L}_{\boldsymbol{X}}(L)$, where $\boldsymbol{X} = \times(\boldsymbol{X}_j \mid j \in \mathbf{J}(L))$, we have

$$\mathbf{L}_{\boldsymbol{X}}(L) = \mathbf{L}_{\boldsymbol{X}}((\mathbf{J}(L), E^{<})) = \mathbf{M}_{\boldsymbol{X}}((\mathbf{J}(L), E^{<})). \tag{4.5}$$

The first equality is established by applying Definition 2.3 to $\mathbf{L}_{\boldsymbol{X}}((\mathbf{J}(L), E^{<}))$, then invoking the identity of $\mathbf{A}((\mathbf{J}(L), E^{<}))$ and $\mathbf{A}((\mathbf{J}(L), \leq))$ and the isomorphism (4.1); and by applying Definition 4.1 (using (4.3)) to $\mathbf{L}_{\boldsymbol{X}}(L)$, then invoking (4.2). The second equality follows immediately from Theorem 3.1. Thus, the LCI model $\mathbf{L}_{\boldsymbol{X}}(L)$ can be expressed as the TADG model $\mathbf{M}_{\boldsymbol{X}}((\mathbf{J}(L), E^{<}))$.

Conversely, consider a TADG model $\mathbf{M}_{\boldsymbol{X}}(D)$, where $D \equiv (V, E)$ is a TADG and where $\boldsymbol{X} = \times(\boldsymbol{X}_a \mid a \in V)$. Then

$$\mathbf{M}_{\boldsymbol{X}}(D) = \mathbf{L}_{\boldsymbol{X}}(D) = \mathbf{L}_{\boldsymbol{X}}(\mathbf{A}(D)); \tag{4.6}$$

the first equality follows from Theorem 3.1, while the second follows from Definitions 2.3 and 4.1 (using (4.4)). Thus, $\mathbf{M}_{\boldsymbol{X}}(D)$ can be expressed as the LCI model $\mathbf{L}_{\boldsymbol{X}}(\mathbf{A}(D))$.

**Remark 4.1.** In general $|\mathbf{J}(L)|$ is smaller that $|L|$, so in (4.5), the TADG representation is more economical than the LCI representation.

**Remark 4.2.** By Theorem 4.1, LCI models inherit all properties of ADG models, in particular the recursive density factorization that, for multivariate normal or multinomial distributions, allows simple and explicit (non-iterative) statistical analysis. As noted earlier, however, LCI models comprise a distinguished subclass of ADG models. Besides the features of LCI models noted in Section 1, we add here that in the normal case, an LCI model remains invariant under the action of a subgroup (determined by the particular lattice) of the group of block-triangular matrices. This invariance leads to an integral representation of classical form for the distribution of the maximal invariant statistic (= generalized eigenvalues) and to the distribution of the likelihood ratio statistic for testing one LCI model against another. See [AP] (1993, 1995a) for details.

## 5. Markov Equivalence of Acyclic Digraphs; the Essential Graph $D^*$.

As noted in Section 1, different ADGs can determine the same graphical Markov model.

**Definition 5.1.** Two ADGs $D_1$ and $D_2$ are *Markov equivalent* on a product space $\boldsymbol{X}$ indexed by $V$ if $\mathbf{M}_{\boldsymbol{X}}(D_1) = \mathbf{M}_{\boldsymbol{X}}(D_2)$. If $D_1$ and $D_2$ are Markov equivalent on *every* such product space $\boldsymbol{X}$, then $D_1$ and $D_2$ are called *Markov equivalent* and we write $D_1 \sim D_2$. The Markov equivalence class containing $D$ is denoted by $[D]$.

A well-known graph-theoretic criterion for the Markov equivalence of ADGs is stated in Theorem 5.1. This criterion was first proved by Verma and Pearl (1992), Corollary 3.2, - also see [AMP] (1996b), Theorem 2.1 - and independently by Frydenberg (1990), Theorem 5.6, for the more general class of chain graphs - also see [AMP] (1996a), Theorem 3.1.

**Theorem 5.1.** Two ADGs are Markov equivalent if and only if they have the same skeleton and the same immoralities (see Figure 5.1).

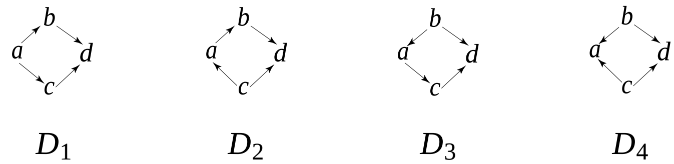



Figure 5.1: The four ADGs with the same skeleton as $D_1$ and the immorality ($b$, $d$, $c$). The ADGs $D_1$, $D_2$, and $D_3$ have no other immoralities, hence are Markov equivalent by Theorem 5.1. The ADG $D_4$ has the additional immorality (b, a, c), hence is not Markov equivalent to the others. Thus, $[D_1] = \{D_1, D_2, D_3\}$.

Since an ADG with $n$ vertices can have at most $O(n^3)$ immoralities, Theorem 5.1 provides a feasible criterion for deciding whether two given ADGs are Markov equivalent in polynomial time. However, it does not directly yield a characterization of the *entire equivalence class* $[D]$ for a given ADG $D$, hence does not provide a feasible criterion for deciding *whether* $[D]$ *contains a transitive* ADG, i.e., *whether $D$ is Markov equivalent to some LCI model.*

Consider, for example, the non-transitive ADG $D_1$ in Figure 5.2: does $[D_1]$ contain a transitive ADG? Theorem 5.1 does not allow us to answer this question by direct inspection of $D_1$; instead, we must first determine all members of $[D_1]$ as follows, then check each member for transitivity. Since ($b$, $d$, $c$) is an immorality in $D_1$, the arrows $b \to d$ and $c \to d$ are *essential* in $D_1$, i.e., these arrows must occur in every member of $[D_1]$. The remaining three edges of $D_1$ might be oriented in $2^3 = 8$ possible ways, as shown in Figure 5.2. Of these 8 digraphs, only 5 are acyclic, and of these 5, only three ($D_1$, $D_2$, $D_3$) possess exactly the same immorality as $D_1$, hence $[D_1] = \{D_1, D_2, D_3\}$. Thus $[D_1]$ does contain a transitive ADG, namely $D_3$.

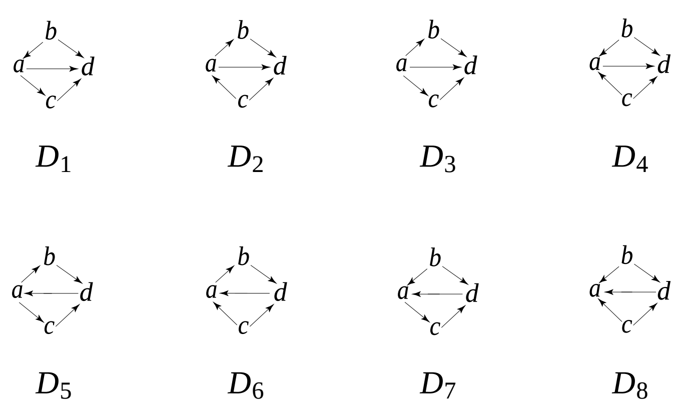


Figure 5.2: The $2^3 = 8$ possible digraphs with the same skeleton as $D_1$ and the immorality $(b, d, c)$. Of these 8, $D_5$, $D_6$, and $D_7$ are not acyclic, while $D_4$ and $D_8$ are acyclic but possess the additional immorality $(b, a, c)$, so $[D_1] = \{D_1, D_2, D_3\}$.

By contrast, consider the non-transitive ADG $D$: $a \to b \to c \to d$. Of the $2^3 = 8$ ADGs with the same skeleton as $D$, it is straightforward to verify that exactly four have one immorality, hence are not Markov equivalent to $D$, while the remaining four (including $D$ itself) have no immoralities and thus comprise the equivalence class $[D]$. Furthermore, no member of $[D]$ is transitive, hence the ADG model determined by $D$ is *not* Markov equivalent to any LCI model.

Since the number of possible orientations of all arrows that do not participate in any immorality of an ADG $D$ grows exponentially with the number of such arrows, hence super-exponentially with the number of vertices, determination of the equivalence class $[D]$ by exhaustive enumeration of possibilities rapidly becomes computationally infeasible as the size of $D$ increases. A closer examination of the ADGs in Figure 5.2 reveals, however, that the arrow $a \to d$ occurs in every member of $[D_1]$, hence is an essential arrow of $D_1$ even though it is not involved in any immorality of $D_1$. Had we been able to identify all 3 essential arrows of $D_1$ directly from $D_1$ itself, it would not have been necessary to consider $D_5$ - $D_8$ in order to determine $[D_1]$. On the other hand, it appears necessary to determine $[D_1]$ before we can identify the essential arrows of $D_1$.

Fortunately, this is not the case. [AMP] (1996b) obtain a computationally feasible polynomial-time algorithm for determining all essential arrows of an ADG $D$ (Theorem 5.3 below). This is done by first introducing and characterizing the *essential graph* $D^*$ associated with $D$ (see Theorem 5.2). Questions such as the existence of a transitive member of $[D]$ can be answered by a polynomial-time inspection of $D^*$ itself (§6), without the need for an exhaustive search of $[D]$.

**Definition 5.2.** The *essential graph* $D^*$ associated with $D$ is the graph

$$D^* := \cup(D' \mid D' \sim D),$$

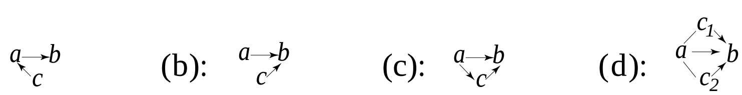
i.e., $D^*$ is the smallest graph larger than every $D' \in [D]$. (See Appendix for definitions.)

Thus, $D^*$ is the graph with the same skeleton as $D$, but where *an edge is directed in* $D^*$ *iff it occurs as a directed edge (≡ arrow) with the same orientation in every* $D' \in [D]$; all other edges of D* are undirected. (See Figure 5.3 for examples.) The directed edges (≡

arrows) in $D^*$ are called the *essential arrows* of $D$. Clearly, every arrow that participates in an immorality in $D$ is essential, but $D$ may contain other essential arrows as well, e.g., the arrow $a \to d$ in the second graph in Figure 5.3 and the arrows $a \to d$ and $b \to d$ (verify!) in the third graph in Figure 5.3. [AMP] (1996b) show that $D^*$ is a chain graph that is itself Markov equivalent to $D$, so that $D^*$ contains the same statistical information as $D$. Their complete characterization of essential graphs, stated here as Theorem 5.2, involves further restrictions on the configurations of arrows and lines (≡ undirected edges) that can occur in $D^*$.


Figure 5.3: Four examples of essential graphs $D^*$. In the first example, $D$ is the ADG $D_1$ of Figure 5.1. In the second, $D$ is the ADG $D_1$ of Figure 5.2. In the third, $D = D^*$ (see Corollary 5.2). In the fourth, $D$ is the ADG $a \to b \to c \to d$.

**Definition 5.3.** Let $G$ be a graph. An arrow $a \to b \in G$ is *strongly protected* in $G$ if $a \to b$ occurs in at least one of the following four configurations as an induced subgraph of $G$:




($c_1 \neq c_2$).

**Theorem 5.2 (Characterization of $D^*$).** A graph $G \equiv (V, E)$ is equal to $D^*$ for some ADG $D$ if and only if $G$ satisfies the following four conditions:

(i) $G$ is a chain graph;

(ii) for every chain component $\tau$ of $G$, $G_\tau$ is chordal;

(iii) the configuration $a \to b — c$ does not occur as an induced subgraph of $G$;

(iv) every arrow $a \to b \in G$ is strongly protected in $G$.

Since both UGs and ADGs are chain graphs, Theorem 5.2 immediately yields the following two corollaries.

**Corollary 5.1.** Let $G$ be a UG. Then $G = D^*$ for some ADG $D$ if and only if $G$ is chordal.

**Corollary 5.2.** Let $G$ be a digraph. Then $G = D^*$ for some ADG $D$ if and only if $G$ is an ADG and every arrow of $G$ is strongly protected in $G$ (note that configuration (d) cannot occur since $G$ has no undirected edges). In this case $G = D = D^*$.

The following results will be applied in Section 6.

**Proposition 5.1** (≡ Proposition 4.2 of [AMP] (1996b)). A digraph $D'$ is acyclic and equivalent to the ADG $D$ if and only if $D'$ is obtained from $D^*$ by orienting the edges of each (chordal) chain component $(D^*)_\tau$ in any perfect (≡ acyclic and moral) way.

**Proposition 5.2** (≡ Proposition 4.3 of [AMP] (1996b)). An ADG $D$ is Markov equivalent[5] to its essential graph $D^*$.

[AMP] (1996b) presented and verified the following polynomial-time algorithm for constructing the essential graph $D^*$ associated with an ADG $D$.

**Theorem 5.3 (the Construction Algorithm).** Define $G_0 := D$. For $i \geq 1$, convert every arrow $a \rightarrow b \in G_{i-1}$ that is *not* strongly protected in $G_{i-1}$ into a line $a — b$, obtaining a graph $G_i$. Stop after $k$ steps, where $k \geq 0$ is the smallest nonnegative integer such that $G_k = G_{k+1}$. Necessarily, $k \leq |E|$. Then $G_k = D^*$.

This algorithm produces a sequence $G_0, \cdots, G_k$ of graphs such that $D \equiv G_0 \subset \cdots \subset G_k = G_{k+1}$. Since both arrows of an immorality are strongly protected, each $G_i$ has the same immoralities as $D$ and $D^*$. Let $n = |V|$. Because the determination of the set of arrows that are not strongly protected in $G_{i-1}$ requires at most $O(n^4)$ operations, this algorithm requires at most $O(|E|\,n^4)$ operations, although it can be implemented in a more efficient fashion (cf. Chickering (1995), Meek (1995)).

## 6. A Graphical Characterization of LCI Models.

The following characterization theorem leads to a feasible (≡ polynomial-time) algorithm for deciding whether an ADG $D$ is Markov equivalent to some LCI model:

**Theorem 6.1.** Let $D \equiv (V, E)$ be an ADG. The Markov-equivalence class $[D]$ contains at least one transitive ADG (equivalently, $D$ is Markov equivalent to some LCI model) if

[5]For the definition of Markov equivalence of chain graphs, of which UGs, ADGs, and essential graphs are special cases, see for example Appendix B of Andersson *et al* (1995b).

and only if none of the following five configurations occurs as an induced subgraph of the associated essential graph $D^*$:

($\alpha$): $a \to b \to c$ ($\beta$): $\begin{array}{ccc} a & \to & b \\ | & & \uparrow \\ c & — & d \end{array}$ ($\gamma$): $a \leftarrow b — c \to d$

($\delta$): $a — b — c \to d$ ($\varepsilon$): $a — b — c — d$

The proof of Theorem 6.1 is accomplished by means of the following algorithm for orienting the undirected edges of a general essential graph $D^* \equiv (V, E^*)$.

**The Orientation Algorithm.** For each vertex $a \in V$, define the *degree* $\nu(a)$ of $a$ in $D^*$ as follows:

$$\nu(a) := |\{a' \in V \mid a — a' \in D^* \text{ or } a \to a' \in D^*\}|.$$

For each chain component $\tau \in \mathbf{T}(D^*)$ such that $|\tau| \geq 2$, orient the (necessarily undirected) edges of $(D^*)_\tau$ as follows:

Step 1.1: Define $a_1 \equiv a_1(\tau) := \text{argmax}\{\nu(a) \mid a \in \tau\}$.

Step 1.2: Define $\tau_1 \equiv \tau_1(\tau) := \{a' \in \tau \mid a_1 — a' \in (D^*)_\tau\} \cup \{a_1\}$.

Step 1.3: $\forall a' \in \tau_1 \setminus \{a_1\}$, orient the edge $a_1 — a'$ as $a_1 \to a'$.

For $i \geq 2$:

Step $i$.1: Define $a_i \equiv a_i(\tau) := \text{argmax}\{\nu(a) \mid a \in (\tau_1 \cup \ldots \cup \tau_{i-1}) \setminus \{a_1, \ldots, a_{i-1}\}\}$.

Step $i$.2: Define $\tau_i \equiv \tau_i(\tau) := \{a' \in \tau \mid a_i — a' \in (D^*)_\tau\} \setminus \{a_1, \ldots, a_{i-1}\}$.

Step $i$.3: $\forall a' \in \tau_i$, orient the edge $a_i — a'$ as $a_i \to a'$.

Terminate the procedure after Step $k$, where $k := \min\{i \mid \{a_1, \ldots, a_i\} = \tau_1 \cup \ldots \cup \tau_i\}$. Clearly, $a_1, \ldots, a_k$ are distinct vertices in $\tau$, $a_1 \in \tau_1$, and $a_i \in \tau_1 \cup \ldots \cup \tau_{i-1} \subseteq \tau$ for $2 \leq i \leq k$, so $k \leq |\tau|$.

**Fact 1.** $k = |\tau|$.
**Proof.** If $k < |\tau|$, then $\sigma := \tau \setminus \{a_1, \ldots, a_k\} \neq \emptyset$. Since $(D^*)_\tau$ is connected, there exists $a \in \sigma$ and $a_i \in \{a_1, \ldots, a_k\}$ such that $a — a_i \in (D^*)_\tau$. But this implies that $a \in \tau_i \subseteq \tau_1 \cup \ldots \cup \tau_k = \{a_1,$

…, $a_k\}$, contradicting the assumption that $a \in \sigma$.

In view of Fact 1, $\{a_1, \ldots, a_k\} = \tau_1 \cup \ldots \cup \tau_k = \tau$ and all edges in $(D^*)_\tau$ are oriented by this procedure. Therefore, by applying this Orientation Algorithm to every chain component $\tau \in \mathbf{T}(D^*)$ we obtain a digraph $D' \subseteq D^*$ with the same skeleton as $D^*$. If $a_j \to a_i \in (D')_\tau$, then necessarily $j < i$. Also, for each $a_i \in \{a_2, \ldots, a_k\}$, $a_i \in (\tau_1 \cup \ldots \cup \tau_{i-1}) \setminus \{a_1, \ldots, a_{i-1}\}$, so there exists $j < i$ such that $a_j \to a_i \in (D')_\tau$. Therefore, for each $a_i \in \{a_2, \ldots, a_k\}$ there exists a directed path from $a_1$ to $a_i$ in $(D')_\tau$.

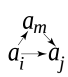
**Fact 2.** The digraph $(D')_\tau$ is acyclic.

**Proof.** This is immediate from the first and third sentences of the preceding paragraph.

**Fact 3.** The ADG $(D')_\tau$ is moral.

**Proof.** Suppose that an immorality $a_i \to a_m \leftarrow a_j$ occurs in $(D')_\tau$. By the last sentence of the paragraph preceding Fact 2, there exist directed paths from $a_1$ to $a_i$ and from $a_1$ to $a_j$ in $(D')_\tau$. Since $i, j < m$, neither path contains $a_m$. Because $a_i$ and $a_j$ are not adjacent in $(D')_\tau$, this contradicts the fact that $(D^*)_\tau$ is chordal (Theorem 5.2(ii)).

**Fact 4.** (i) $\nu(a_1) \geq \nu(a_i)$ for $i \geq 2$.

(ii) If the configuration 

occurs in $(D')_\tau$, then $\nu(a_m) \geq \nu(a_j)$.

**Proof.** (i) is obvious. For (ii), by the definition of $a_m$ it suffices to show that $a_j \in (\tau_1 \cup \ldots \cup \tau_{m-1}) \setminus \{a_1, \ldots, a_{m-1}\}$. This is a consequence of the following three observations:

(a) $a_i \to a_j \in (D')_\tau \Rightarrow a_j \in \tau_i$;

(b) $a_i \to a_m \in (D')_\tau \Rightarrow i < m \Rightarrow \tau_i \subseteq \tau_1 \cup \ldots \cup \tau_{m-1}$;

(c) $a_m \to a_j \in (D')_\tau \Rightarrow m < j \Rightarrow a_j \notin \{a_1, \ldots, a_{m-1}\}$.

**Proposition 6.1.** Let $D$ be an ADG and let $D'$ be the digraph obtained by applying the Orientation Algorithm to the associated essential graph $D^*$. Then $D'$ is acyclic and equivalent to $D$.

**Proof.** This is an immediate consequence of Facts 1-3 and Proposition 5.1 .

**Proof of Theorem 6.1.** If any of the five configurations ($\alpha$) -($\varepsilon$) occur as an induced subgraph of $D^*$, then every orientation of the undirected edges of $D^*$ must produce a non-transitive digraph, hence every ADG in [*D*] is non-transitive.

Conversely, if none of these configurations occur in *D**, we shall show that the equivalent ADG *D′* obtained by applying the Orientation Algorithm to *D** is transitive. Suppose to the contrary that *D′* is non-transitive. For the purposes of this proof only, we shall denote the essential arrows of *D′* (i.e., the arrows of *D**) by —» and the non-essential arrows of *D′* (i.e., those obtained from lines of *D** by the Orientation Algorithm) by —›. Thus, at least one of the following four non-transitive configurations must occur as an induced subgraph of *D′*:

(1): *a* »*b* »*c*     (2): *a* »*b* ›*c*     (3): *a* ›*b* »*c*     (4): *a* ›*b* ›*c*

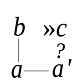
Clearly (1) cannot occur in *D′* since this would violate the non-occurrence of (α) in *D**, while (2) cannot occur in *D′* since *D** would fail to satisfy condition (iii) of Theorem 5.2.

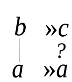
Suppose that (3) occurs in *D′*. We assert that $A \setminus B = \emptyset$, where

$$A := \{a' \in V \mid a—a' \in D^* \text{ or } a—\!\!\gg a' \in D^*\},$$

$$B := \{a' \in V \mid b—a' \in D^* \text{ or } b—\!\!\gg a' \in D^*\}.$$

If $A \setminus B \neq \emptyset$, then each $a' \in A \setminus B$ must occur in one of the following two configurations as an induced subgraph of *D**:

(I): 

,     (II): 

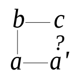
,

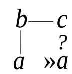
where "?" indicates either no edge, an undirected edge, or a directed edge of unspecified orientation. (The occurrence of an arrow *b*«—*a′* in (I) or (II) would induce a directed cycle in *D**, hence is forbidden.) The absence of any edge between *a′* and *c* in (I) or (II) would contradict the non-occurrence in *D** of (δ) and (γ), respectively, while the occurrence of *a′*—*c* in (I) or (II) would contradict condition (iii) of Theorem 5.2. The occurrence of *a′*—»*c* in (I) or (II) would contradict the non-occurrence in *D** of (β) and (α), respectively. Finally, the occurrence of *c*—»*a′* in (I) or (II) would contradict the non-occurrence in *D** of (α). Thus $A \setminus B = \emptyset$.

Because $c \in B \setminus A$, it follows that $\nu(b) \equiv |B| > |A| \equiv \nu(a)$. Since $a—b \in D^*$, $\{a, b\} \subseteq \tau$ for some $\tau \in \mathbf{T}(D^*)$. If $a = a_1 \equiv a_1(\tau)$ then $\nu(a) \geq \nu(b)$ by Fact 4(i), a contradiction. If $a = a_i$

$\equiv a_i(\tau)$ for some $i \geq 2$, then there exists $j < i$ such that $a \leftarrow a_j \in (D')_\tau$. Therefore $a_j \in A \subseteq B$, hence either $b \leftarrow a_j \in (D')_\tau$, $b \rightarrow a_j \in (D')_\tau$, or $b \twoheadrightarrow a_j \in (D')_\tau$. Since $a \rightarrow b \in (D')_\tau$, the second and third possibilities would violate the acyclicity of $D'$, hence the first possibility must hold. By Fact 4(ii), this implies that $\nu(a) \geq \nu(b)$, again a contradiction. Thus (3) cannot occur in $D'$.

Lastly, suppose that (4) occurs in $D'$. We again assert that $A \setminus B = \emptyset$. If $A \setminus B \neq \emptyset$, then as above, each $a' \in A \setminus B$ must occur in one of the following two configurations as an induced subgraph of $D^*$:

(III): 

(IV): 


The absence of any edge between $a'$ and $c$ in these two configurations would contradict the non-occurrence in $D^*$ of ($\varepsilon$) and ($\delta$), respectively, while the occurrence of $a'$—$c$ would contradict conditions (ii) and (iii), respectively, of Theorem 5.2. The occurrence of $a' \twoheadrightarrow c$ in (III) and (IV) would contradict both the acyclicity of $D^*$ and condition (iii). Finally, the occurrence of $c \twoheadrightarrow a'$ in (III) would contradict both acyclicity and condition (iii), while its occurrence in (IV) would contradict the non-occurrence of ($\beta$) in $D^*$. Thus, again $A \setminus B = \emptyset$.

Because $c \in B \setminus A$, again $\nu(b) > \nu(a)$. Here, $\{a, b, c\} \subseteq \tau$ for some $\tau \in \mathbf{T}(D^*)$. Exactly as above, however, we can show that $\nu(a) \geq \nu(b)$, a contradiction. Thus (4) cannot occur in $D'$. This completes the proof of Theorem 6.1.

The first essential graph $D^*$ in Figure 5.3 coincides with configuration ($\beta$), the third contains ($\alpha$) as an induced subgraph, while the fourth coincides with ($\varepsilon$), hence Theorem 6.1 implies that none of the three associated ADGs is Markov equivalent to a TADG, so none is Markov equivalent to any LCI model. The second $D^*$ contains none of the configurations ($\alpha$) - ($\varepsilon$) as an induced subgraph, hence its associated ADG *is* Markov equivalent to some TADG, hence to some LCI model. (These facts were established in Section 5 by exhaustive enumeration of [$D$].)

As noted in Section 1, the only undirected graphical (UG) models that share the amenable statistical properties of ADG models are the decomposable models, which can be characterized on the one hand as those determined by *chordal* UGs, and on the other as those UGs that are Markov equivalent to some ADG (Dawid and Lauritzen (1993); [AMP] (1996a), Corollary 4.5). The following corollary characterizes those UGs that are Markov equivalent to some TADG $\equiv$ LCI model.

**Corollary 6.1.** A UG $G$ is Markov equivalent to some LCI model if and only if $G$ is chordal and does not contain configuration (ε): $a$—$b$—$c$—$d$ as an induced subgraph.

**Proof.** ("only if"): By Theorem 4.1, $G$ is Markov equivalent to some TADG $D$, hence $G$ is chordal. By the basic Markov equivalence theorem for chain graphs (Frydenberg (1990), Theorem 5.6; [AMP] (1996a), Theorem 3.1), $D^u = G^u \equiv G$. Therefore, since $D$ is transitive, $G$ cannot contain configuration (ε) as an induced subgraph

("if"): By Corollary 5.1, $G = D^*$ for some ADG $D$. Thus $D^*$ is a UG, hence cannot contain configurations (α) - (δ) as induced subgraphs, and does not contain (ε) by assumption. By Theorem 6.1, therefore, $D$ is Markov equivalent to some LCI model, hence, by Proposition 5.2, so is $D^* \equiv G$.

**Remark 6.1.** It was noted in Section 5 that $D^*$ can be obtained from $D \equiv (V, E)$ by an algorithm that can be implemented in polynomial-time in $n \equiv |V|$. Since each of the configurations (α) - (ε) involves at most four vertices, at most $O(n^4)$ operations are required to determine whether $D^*$ contains any of (α) - (ε). Thus, *it is possible to determine in polynomial-time whether a specific ADG D is Markov equivalent to some TADG and hence to some LCI model*. If it is equivalent, then the Orientation Algorithm produces a Markov-equivalent TADG $D'$ from $D^*$, also in polynomial time (cf. Peyton *et al* (1993)).

## 7. LCI Models Determined by Rings of Subsets.

In [AP] (1991, 1993, 1994, 1995a, b), [AMPT] (1995), and [PW] (1996), LCI models are defined in a slightly more restrictive fashion than in Section 4 above. The finite distributive lattice is presented as a ring[6] $\boldsymbol{K}$ of subsets of a finite index set $I$ and the LCI model consists of a family of multivariate distributions on a product space of the form $\boldsymbol{Y} \equiv \times(\boldsymbol{Y}_i \mid i \in I)$. Such a distribution $P$ is represented by a random variate $Y := (Y_i \mid i \in I) \in \boldsymbol{Y}$. For any subset $K \subseteq I$, define $Y_K := (Y_i \mid i \in K)$, abbreviate $Y_i$ and $Y_K$ by $i$ and $K$, respectively, and define $Y_\varnothing \equiv$ constant. Then the LCI model $\mathbf{L}(\boldsymbol{K}) \equiv \mathbf{L}_{\boldsymbol{Y}}(\boldsymbol{K})$ is defined to be the set of all $P$ such that

$$(7.1) \qquad L \perp M \mid L \cap M \; [P]$$

for every $L, M \in \boldsymbol{K}$.

---

[6]A ring of subsets of $I$ is closed under finite intersections and unions, and contains $I$ and Ø.

In order to express $\mathbf{L}_{\boldsymbol{Y}}(\boldsymbol{K})$ as a general LCI model $\mathbf{L}_{\boldsymbol{X}}(L)$, following [AP] (1993) define, for each $K \in \boldsymbol{K}$,

$$<K> := \cup(K' \in \boldsymbol{K} \mid K' \subset \boldsymbol{K})$$

$$[K] := K \setminus <K>.$$

It is straightforward to verify that

$$\mathbf{J}(\boldsymbol{K}) = \{J \in \boldsymbol{K} \mid <J> \subset J\} \equiv \{J \in \boldsymbol{K} \mid [J] \neq \emptyset\}.$$

If we now define

$$[\mathbf{J}(\boldsymbol{K})] := \{[J] \mid J \in \mathbf{J}(\boldsymbol{K})\},$$

then, since $\mathbf{J}(\boldsymbol{K})$ is a poset under the set inclusion ordering $\subseteq$, $[\mathbf{J}(\boldsymbol{K})]$ is an isomorphic poset under the induced ordering $\leq$ defined as follows: $[J'] \leq [J]$ iff $J' \subseteq J$. Figures 1.2 and 7.1 illustrate the relations among $\boldsymbol{K}$, $(\mathbf{J}(\boldsymbol{K}), \subseteq)$, and $([\mathbf{J}(\boldsymbol{K})], \leq)$.

$\{a, b\}$ $\{a, c\}$
\ /
$\{a\}$

(a) $\mathbf{J}(\boldsymbol{K})$: $\{a\} \subseteq \{a, b\}$, $\{a\} \subseteq \{a, c\}$

$[\{a, b\}] \equiv \{b\}$ $\{c\} \equiv [\{a, c\}]$
\ /
$\{a\}$

(b) $[\mathbf{J}(\boldsymbol{K})]$: $\{a\} \leq \{b\}$, $\{a\} \leq \{c\}$

Figure 7.1. The isomorphic posets $(\mathbf{J}(\boldsymbol{K}), \subseteq)$ and $([\mathbf{J}(\boldsymbol{K})], \leq)$ determined by the ring $\boldsymbol{K}$ in Figure 1.2. Note that $\boldsymbol{K}$ can be recovered from $[\mathbf{J}(\boldsymbol{K})]$ by (7.2).

The ring $\boldsymbol{K}$ can be reconstructed from the members of $[\mathbf{J}(\boldsymbol{K})]$ by the relation

(7.2) $$L = \cup([J] \mid J \in \mathbf{J}(\boldsymbol{K}), J \subseteq L) \in \mathbf{A}(([\mathbf{J}(\boldsymbol{K})], \leq)),$$

valid for all $L \in \boldsymbol{K}$ (cf. [AP] (1993), Proposition 2.1), a special case of the isomorphism (4.1). Thus, each $L \in \boldsymbol{K}$ is represented uniquely as a disjoint union of members of the poset $([\mathbf{J}(\boldsymbol{K})], \leq)$; in particular[7],

(7.3) $$I = \cup([J] \mid J \in \mathbf{J}(\boldsymbol{K})).$$

We can now express $\mathbf{L}_{\boldsymbol{Y}}(\boldsymbol{K})$ as a general LCI model $\mathbf{L}_{\boldsymbol{X}}(L)$: simply set $L = \boldsymbol{K}$ and define $\boldsymbol{X} := \times(\boldsymbol{X}_J \mid J \in \mathbf{J}(\boldsymbol{K}))$, where $\boldsymbol{X}_J := \times(\boldsymbol{Y}_i \mid i \in [J])$. Then $\boldsymbol{X} = \boldsymbol{Y}$ by (7.3), and the LCI condition

[7]See, for example, Figure 7.1(b), where $I = \{a, b, c\}$.

(4.4) for $\mathbf{L}_{\boldsymbol{X}}(L)$ coincides with the LCI condition (7.1) for $\mathbf{L}_{\boldsymbol{Y}}(\boldsymbol{K})$. Thus the LCI model $\mathbf{L}_{\boldsymbol{Y}}(\boldsymbol{K})$ is a general LCI model $\mathbf{L}_{\boldsymbol{X}}(L)$ with the slight restriction that each component space $\boldsymbol{X}_J$ is itself a product space indexed by the members of $[J]$.

## Appendix: Graphs.

Our terminology and notation closely follows those of Lauritzen *et al* (1990) and Frydenberg (1990), with one exception noted below. A *graph* $G$ is a pair $(V, E)$, where $V$ is a finite set of *vertices* and $E$, the set of *edges*, is a subset of $E^*(V) \equiv (V\times V)\setminus\{(a, a) \mid a \in V\}$, i.e., a set of ordered pairs of distinct vertices; thus our graphs include no loops or multiple edges. An edge $(a, b) \in E$ whose opposite $(b, a) \in E$ is called an *undirected* edge and appears as a *line* $a$—$b$ in our figures, whereas an edge $(a, b) \in E$ whose opposite $(b, a) \notin E$ is called a *directed* edge and appears as an *arrow*[8]: $a \rightarrow b$. If $G$ contains only undirected edges, it is an *undirected graph* (UG); if $G$ contains only directed edges it is a *directed graph* (*digraph*).

It shall be convenient to write "$a \rightarrow b \in G$" to indicate that $(a, b) \in E$ but $(b, a) \notin E$; in this case we say that *the arrow* $a \rightarrow b$ *occurs in* $G$. Similarly, we write "$a$—$b \in G$" to indicate that $(a, b) \in E$ and $(b, a) \in E$; in this case we say that *the line* $a$—$b$ *occurs in* $G$.

For each vertex $a \in V$, define $\mathrm{pa}_G(a) := \{b \in V \mid b \rightarrow a \in G\}$, the set of *parents* of $a$ in $G$. For any subset $A \subseteq V$, the *boundary* of $A$ in $G$ is the set $\mathrm{bd}_G(A) := \{b \in V\setminus A \mid (b, a) \in E$ for some $a \in A\}$; the *closure* of $A$ in $G$ is the set $\mathrm{cl}_G(A) := \mathrm{bd}_G(A) \cup A$.

A subset $A \subseteq V$ *induces* the subgraph $G_A := (A, E_A)$, where $E_A := E \cap (A\times A)$.

The *skeleton* $G^u$ of a graph $G \equiv (V, E)$ is its underlying undirected graph, i.e., $G^u := (V, E^u)$, where $E^u := \{(a,b) \mid (a, b) \in E$ or $(b, a) \in E\}$. Two vertices $a$, $b$ are called *adjacent* in $G$ if $(a, b) \in E^u$.

Let $a$, $b$, and $c$ be three distinct vertices of $G \equiv (V, E)$. The triple $(a, b, c)$ is called an *immorality* of $G$ if the induced subgraph $G_{\{a, b, c\}}$ is $a \rightarrow b \leftarrow c$; that is, if the "parents" $a$ and $c$ of $b$ are "unmarried" ($\equiv$ non-adjacent).

A graph $G_2 \equiv (V_2, E_2)$ i s said to be *larger* than a graph $G_1 \equiv (V_1, E_1)$, denoted by $G_1 \subseteq G_2$, if $V_1 \subseteq V_2$ and $E_1 \subseteq E_2$. Thus, if $(G_1)^u = (G_2)^u$, then $G_1 \subseteq G_2$ iff $G_1$ and $G_2$ differ only in that some directed edges (arrows) in $G_1$ may be converted into undirected edges (lines) in $G_2$. We write $G_1 \subset G_2$ if $G_1 \subseteq G_2$ but $G_1 \neq G_2$.

The *union* of a finite collection of subgraphs $(G_i \equiv (V_i, E_i) \mid i = 1, \cdots, n)$ of $G \equiv (V, E)$ is

[8]Our notation differs from Frydenberg's in this regard: he uses the notation $a \Rightarrow b$ rather than $a \rightarrow b$ in his text, although not in his figures.

the subgraph $\cup G_i := (\cup V_i, \cup E_i)$. Clearly, $\cup G_i$ is the smallest subgraph larger than each $G_i$, $i = 1, \cdots, n$.

Let $a$, $b$ be distinct vertices in $G \equiv (V, E)$. A *path* $\pi$ of length $n \geq 1$ from $a$ to $b$ in $G$ is a sequence $\pi \equiv \{a_0, a_1, \cdots, a_n\} \subseteq V$ of distinct vertices such that such that $a_0 = a$, $a_n = b$, and either $a_{i-1} \rightarrow a_i \in G$ or $a_{i-1} — a_i \in G$ for every $i = 1, \cdots, n$. If $a_{i-1} \rightarrow a_i \in G$ for at least one $i$, the path is *directed*; if this is not the case, the path is *undirected*. A (*directed*) *cycle* is a (directed) path with the modification that $a_0 = a_n$. An arrow $a \rightarrow b \in G$ is said to *block a directed cycle in G* if there is a directed path from $a$ to $b$ in $G$ other than $a \rightarrow b$ itself.

A UG $G \equiv (V, E)$ is *complete* if all pairs of vertices are adjacent. Trivially, the empty graph is complete. A subset $A \subseteq V$ is *complete* if its induced subgraph $G_A$ is complete. A subset $A \subseteq V$ is *connected* in $G$ if for every distinct pair $a, b \in A$, there is a path from $a$ to $b$ in $G_A$. For pairwise disjoint subsets $A$ ($\neq \emptyset$), $B$ ($\neq \emptyset$), and $S$ of $V$, $A$ and $B$ are *separated* by $S$ in $G$ if all paths from vertices in $A$ to vertices in $B$ intersect $S$.

The UG $G \equiv (V, E)$ is *chordal* if every cycle of length $n \geq 4$ possesses a *chord*, that is, two non-consecutive adjacent vertices. A total ordering of $V$ is a *perfect ordering* of $G$ if, when each edge of $G$ is oriented in accordance with this ordering, the resulting ADG $D$ is *perfect*, i.e., is *acyclic* and *moral* (without immoralities); $D$ is called a *perfect directed version of* $G$. It is well-known that a UG admits a perfect directed version if and only if it is chordal (cf. Blair and Peyton (1993)).

A graph $G \equiv (V, E)$ is called a *chain graph* if it does not contain any directed cycles. Every induced subgraph $G_A$ of $G$ is also a chain graph. Any UG is trivially a chain graph. A chain graph that is also a digraph is called an *acyclic digraph* (ADG).

For the remainder of the Appendix, let $G \equiv (V, E)$ be a chain graph. Then $G$ determines a pre-ordering $(V, \leq)$ as follows: $a \leq b$ iff $a = b$ or there exists a path from $a$ to $b$ in $G$. A subset $A \subseteq V$ is an *anterior* set if $b \leq a \in A \Rightarrow b \in A$. For any subset $A \subseteq V$, an$(A)$ is the smallest anterior set containing $A$: an$(A) = \{b \in V \mid b \leq a$ for some $a \in A\}$.

If both $a \leq b$ and $b \leq a$ then we write $a \approx b$, which occurs iff $a = b$ or there is an *undirected* path from $a$ to $b$ in $G$. Frydenberg (1990) notes that $\approx$ is an equivalence relation on $V$; we denote the set of equivalence classes in $V$ by $\mathbf{T}(G)$. Equivalently, $\mathbf{T}(G)$ is the set of connected components of the undirected graph obtained from $G$ by removing all directed edges. Each $\tau \in \mathbf{T}(G)$ is called a *chain component* of $G$. A connected UG has only one chain component, while for an ADG, every chain component consists of a single vertex.

We write $a < b$ if there exists a *directed* path from $a$ to $b$. The *future* of a vertex $a \in V$ is the set $\phi(a) := \{b \in V \mid a < b\}$.

A triple $(a, C, b)$ is called a *complex* in $G$ if $C$ is a connected subset of a chain component $\tau \in \mathbf{T}(G)$ and $a$ and $b$ are two non-adjacent vertices in $\mathrm{bd}_G(\tau) \cap \mathrm{bd}_G(C)$. A complex $(a, C, b)$ is called a *minimal complex* in $G$ if no proper subset $C' \subset C$ forms a complex $(a, C', b)$ in $G$. Frydenberg(1990) notes that $(a, C, b)$ is a minimal complex in $G$ iff $G_{C \cup \{a, b\}}$ looks like the chain graph of Figure A.1. An immorality is the special case of a minimal complex where $|C| = 1$.

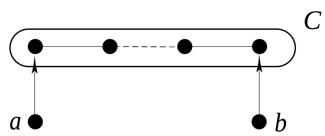



Figure A.1: A simple chain graph. Here $(a, C, b)$ is a minimal complex.

The *moral graph* determined by $G$ is the undirected graph $G^{\mathrm{m}} \equiv (V, E^{\mathrm{m}})$, where $E^{\mathrm{m}} := E^{\mathrm{u}} \cup [\cup(E^*(\mathrm{bd}_G(\tau)) \mid \tau \in \mathbf{T}(G))]$. That is, $G^{\mathrm{m}}$ is $G^{\mathrm{u}}$ augmented by all undirected edges needed to make $\mathrm{bd}_G(\tau)$ complete in $G^{\mathrm{m}}$ for every chain component $\tau \in \mathbf{T}(G)$. Equivalently, $G^{\mathrm{m}}$ is obtained from $G^{\mathrm{u}}$ by adding a line $a — b$ whenever $(a, C, b)$ is a minimal complex in $G$.

## References

Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *J. Amer. Statist. Assoc.* **52** 200-203.

Andersson, S. A. (1990). The lattice structure of orthogonal linear models and orthogonal variance component models. *Scand. J. Statist.* **17** 287-319.

Andersson, S. A., D. Madigan, and M. D. Perlman (1996a). On the Markov equivalence of chain graphs, undirected graphs, and acyclic digraphs. *Scand. J. Statist.*, in press.

Andersson, S. A., D. Madigan, and M. D. Perlman (1996b). A characterization of Markov equivalence classes for acyclic digraphs. *Ann. Statist.*, in press.

Andersson, S. A., D. Madigan, M. D. Perlman, and C. M. Triggs (1995). On the relation between conditional independence models determined by finite distributive lattices and by directed acyclic graphs. *J. Statist. Planning and Inference* **48** 25-46.

Andersson, S. A. and M. D. Perlman (1988). Lattice models for conditional independence in a multivariate normal distribution. Preprint, Institute of Mathematical Statistics, University of Copenhagen and Department of Statistics, University of Washington.

Andersson, S. A. and M. D. Perlman (1991). Lattice-ordered conditional independence models for missing data. *Statist. Probability Letters* **12** 465-486.

Andersson, S. A. and M. D. Perlman (1993). Lattice models for conditional independence in a multivariate normal distribution. *Ann. Statist.* **21** 1318-1358.

Andersson, S. A. and M. D. Perlman (1994). Normal linear models with lattice conditional independence restrictions. *Multivariate Analysis and Its Applications* (T. W. Anderson, K.-T. Fang, and I. Olkin, eds.) IMS Lecture Notes-Monograph Series, Vol. 24, 97-110.

Andersson, S. A. and M. D. Perlman (1995a). Testing lattice conditional independence models. *J. Multivariate Analysis* **53** 18-38.

Andersson, S. A. and M. D. Perlman (1995b). Unbiasedness of the likelihood ratio test for lattice conditional independence models. *J. Multivariate Analysis* **53** 1-17.

Andersson, S. A. and M. D. Perlman (1996). Normal linear regression models with recursive graphical Markov structure. Submitted to *J. Multivariate Analysis.*

Blair, J. R. S. and B. Peyton (1993). An introduction to chordal graphs and clique trees. In *Graph Theory and Sparse Matrix Computation* (A. George, J. R. Gilbert, and J. W. H. Liu, eds.), 1-29. Springer-Verlag, New York.

Chickering, D. M. (1995). A transformational characterization of equivalent Bayesian network structures. In *Uncertainty in Artificial Intelligence: Proceedings of the Eleventh Conference* (P. Besnard, S. Hanks, eds.), 87-98. Morgan Kaufmann, San Francisco.

Cox, D. R. and Wermuth, N. (1996). *Multivariate Dependencies: Models, Analysis, and Interpretation.* Chapman and Hall, London.

Davey, B. A. and H. A. Priestley (1990). *Introduction to Lattices and Order.* Cambridge U. Press, Cambridge, U.K.

Dawid, A. P. and S. L. Lauritzen (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21** 1272-1317.

Edwards, D. (1995). *Introduction to Graphical Modeling.* Springer, New York.

Frydenberg, M. (1990). The chain graph Markov property. *Scand. J. Statist.* **17** 333-353.

Jensen, F. V. (1996). *An Introduction to Bayesian Networks.* U. College London Press, London.

Koster, J. A. (1996). Markov properties of non-recursive causal models. *Ann. Statist.*, in press.

Lauritzen, S. L. (1996). *Graphical Models.* Oxford University Press, Oxford.

Lauritzen, S. L., A. P. Dawid, B. N. Larsen, and H.-G. Leimer (1990). Independence properties of directed Markov fields. *Networks* **20** 491-505.

Lauritzen, S. L. and D. J. Spiegelhalter (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *J. Royal Statist. Soc. (Series B)* **50** 157-224.

Madigan, D. and A. E. Raftery (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *J. Amer. Statist. Assoc.* **89** 1535-1546.

Meek, C. (1995). Causal inference and causal explanation with background knowledge. In *Uncertainty in Artificial Intelligence: Proceedings of the Eleventh Conference* (P. Besnard, S. Hanks, eds.), 403-410. Morgan Kaufmann, San Francisco.

Neapolitan, R. E. (1990). *Probabilistic Reasoning in Expert Systems.* Wiley, New York.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Mateo, CA.

Perlman, M. D. and L. Wu (1996). Lattice conditional independence models for missing data in contingency tables. In preparation.

Peyton, B.W., A. Pothen, and X. Yuan (1993). Partitioning a chordal graph into transitive subgraphs for parallel sparse triangular solution. *Linear Algebra Appl.* **192** 329-354.

Spiegelhalter, D. J., A. P. Dawid, S. L. Lauritzen, and R. G. Cowell (1993). Bayesian analysis in expert systems (with discussion). *Statist. Science* **8** 219-283.

Spiegelhalter, D. J. and S. L. Lauritzen (1990). Sequential updating of conditional probabilities on directed graphical structures. *Networks* **20** 579-605.

Spirtes, P., Glymour, C;., and Scheines, R. (1993). *Causation, Prediction, and Search.* Springer, New York.

Verma, T. and J. Pearl (1990). Equivalence and synthesis of causal models. In *Uncertainty in Artificial Intelligence: Proceedings of the Sixth Conference* (M. Henrion *et al* eds.) 220-227. North Holland.

Verma, T. and J. Pearl (1992). An algorithm for deciding if a set of observed independencies has a causal explanation. In *Uncertainty in Artificial Intelligence: Proceedings of the Eighth Conference* (D. Dubois *et al* eds.), San Francisco: Morgan Kaufman, pp. 323-330.

Wermuth, N. and Lauritzen, N. (1983). Graphical and recursive models for contingency tables. *Biometrika* **70** 537-552.

Whittaker, J. L. (1990). *Graphical Models in Applied Multivariate Statistics.* Wiley, New York.

Wright, S. (1921). Correlation and causation. *J. Agricultural Research* **20** 557-585.

York, J., D. Madigan, I. Heuch, and R. T. Lie (1995). Estimating a proportion of birth defects by double sampling: a Bayesian approach incorporating covariates and model uncertainty. *Applied Statistics* **44** 227-242.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression equiations and tests for aggregation bias. *J. Amer. Statist. Assoc.* **57** 348-368.